Semantic TRIZ™

Mikhail Verbitsky, Ph.D., Sc.D.

Director, Software Applications and Training Worldwide
Invention Machine Corporation
133 Portland Street, Boston, MA 02114
e-mail: verb@invention-machine.com

1. Introduction: Postulates of TRIZ and their implementation in software.

Theory of Inventive Problem Solving, better known in the engineering world by its Russian abbreviation TRIZ, has been widely acknowledged as an effective tool to solve technical problems, especially at the conceptual design stage. Its attractiveness is based on few appealing observations. First, it was noticed, that a significant number of engineering problems can be described in a contradiction template. These contradictions are continuously recurring across different industries, and these same contradictions have frequently been solved using repeating approaches. It has also been discovered that the patterns of technical systems evolution are repeated across different industries as well: specifically, systems are being developed in the directions of increased ideality; harmonization, controllability, and degree of freedom. And finally, it has been broadly recognized that the best innovative solutions seem to adopt scientific effects from diverse fields of science and engineering.

Extrapolation of these observations to a specific problem created a number of tools used by TRIZ practitioners. Matrix of Contradictions is probably the most popular one. When engineer formulates his problem as a contradiction, the Contradiction Matrix helps him to identify representative contradiction and classic approaches to solve this contradiction, the so-called Inventive Principles. This process is based on a simple assumption, that if these Principles demonstrated their effectiveness in similar circumstances in the past (for the same contradiction), they

APPENDIX B

might be helpful for the current situation as well. In the software media [1], the Contradiction Matrix is implemented as is illustrated in Figure 1.

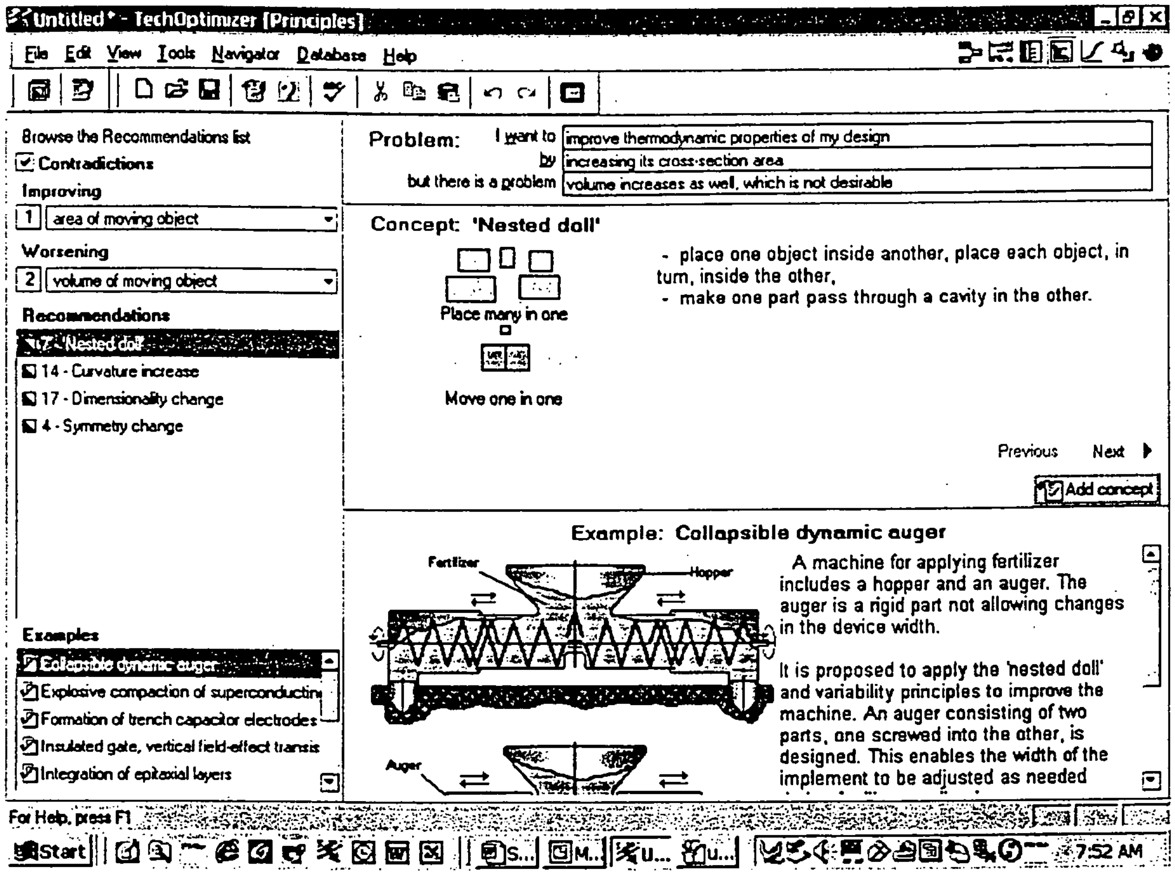


Figure 1: Contradiction Matrix in Invention Machine TechOptimizer [1].

Figure 1 shows the situation, described in the following statement: 'I want to improve thermodynamic properties of my design by increasing its cross-section area, but there are undesirable consequences of the area increase – the volume increases as well'. Matrix of contradictions helps to translate this statement into a contradiction template: improving aspect – area of moving object, worsening aspect – volume of moving object; and suggests several Inventive Principles which might be helpful for this problem because they had demonstrated their effectiveness in similar situations in the past.

Similarly, the distinctive trends of technology evolution have been incorporated into a comprehensive Prediction Tree [1]. There are different techniques to solve engineering problems with its help [2] and one of them is illustrated in Figure 2.

APPENDIX B

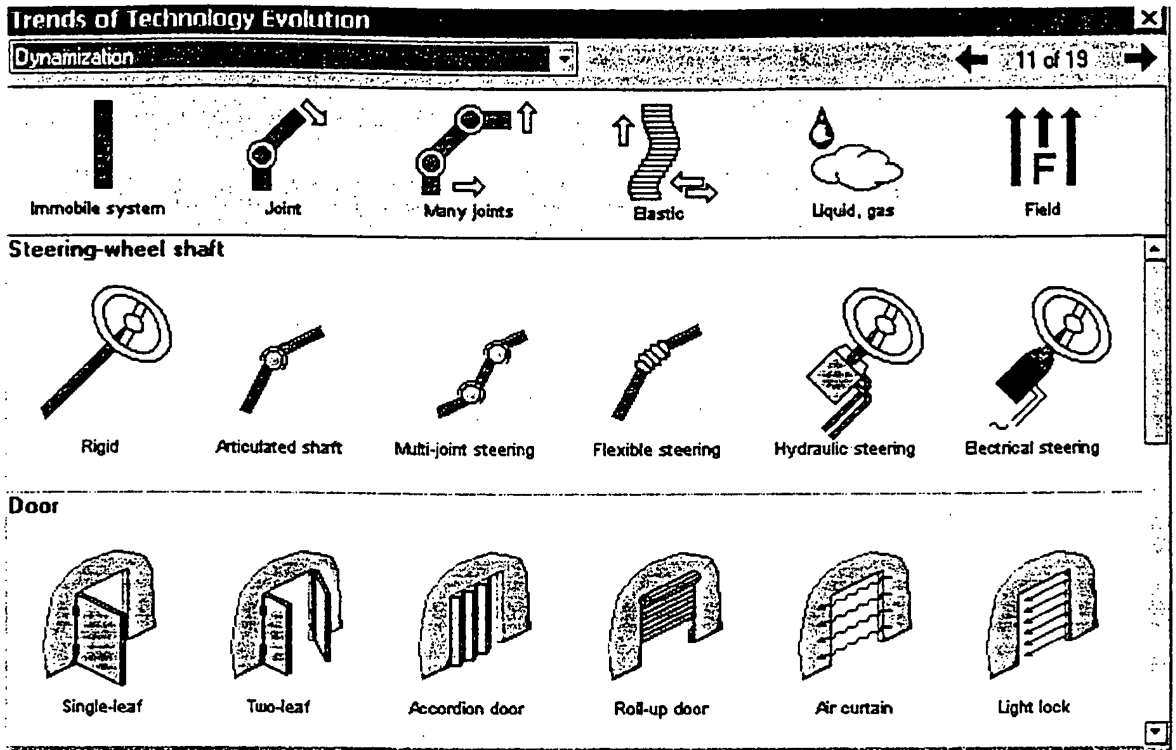


Figure 2: Dynamization trend of engineering systems evolution as described by Invention Machine TechOptimizer [1]

Figure 2 illustrates one of technologies' evolution trends, the trend of Increased Dynamization: engineering systems generally evolve from rigid immobile systems in the direction of increased dynamization, i.e., they employ more joints, increase component's elasticity, replace solid materials with liquid or gas, *et al.* Although the solutions presented by this tool are abstracted templates of what an engineer ultimately implements, their universality serve as a stimulus for generating innovative problem-solving ideas, and lead to the conception of new system features that can improve its performance. The implicit extrapolation assumption behind this tool is similar to that of Inventive Principles: other engineering systems experienced this trend, therefore it is not unlikely that the system we are currently working with may experience the same trend.

The TRIZ idea to search for innovative solutions in different fields of science and engineering has inspired the creation of Scientific Effects knowledge base [1], which is currently the most comprehensive library of its kind available in the world. With about 9000 unique Effects, this library is representative of diverse industries and has nearly nine times the content of the largest scientific encyclopedias. The organization of the Effects knowledge base is suited for problem solvers since it uses a function-oriented taxonomy. Engineers can search this knowledge base by specifying the function they want to perform (rather than topics), and retrieve the scientific phenomena that can be employed to accomplish this task. This process is illustrated by Figure 3.

APPENDIX B

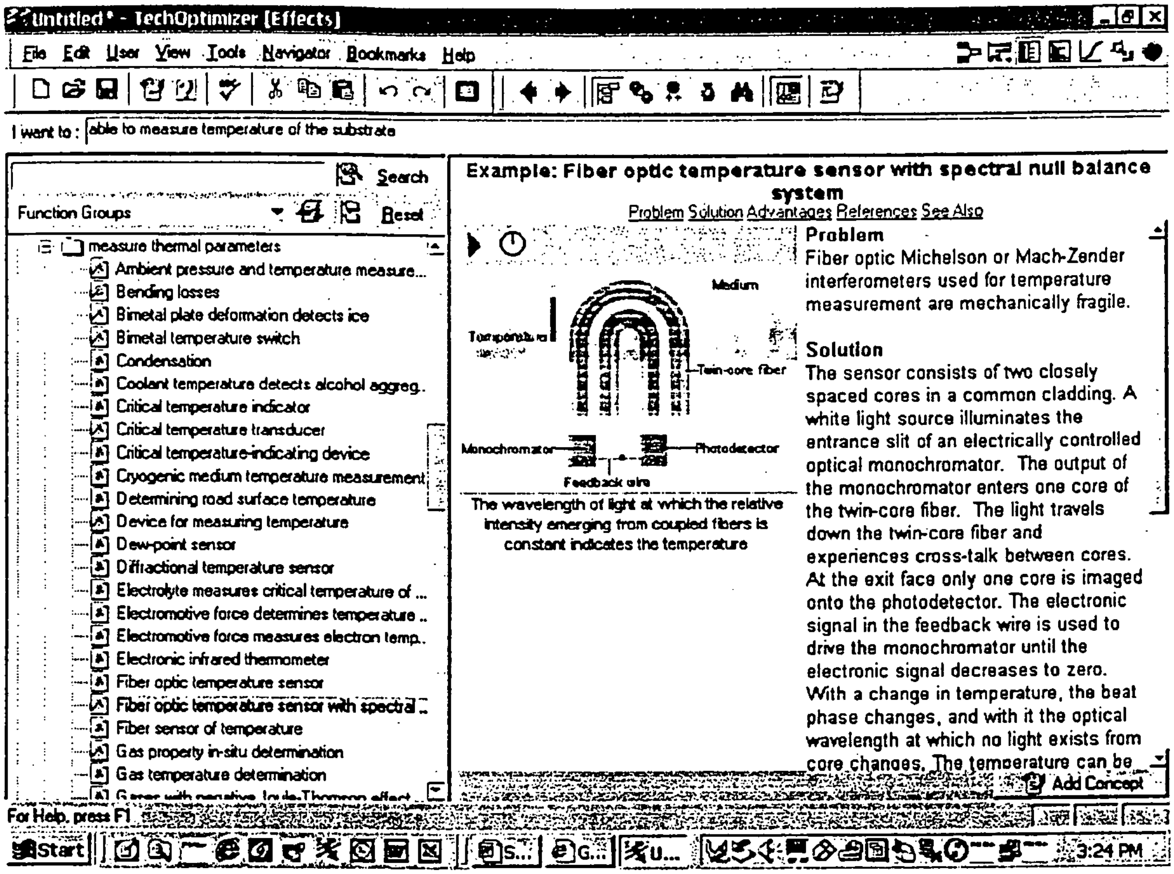


Figure 3: Effects database of Invention Machine TechOptimizer [1].

It demonstrates the situation described in the following statement: 'I want to be able to measure the temperature of the substrate'. Instead of making a decision as to what branch of science should be approached to solve this problem, an engineer can simply open a folder the name of which can be associated with the problem formulated above, i.e., 'measure thermal parameters', and review a variety of scientific effects which can measure temperature.

2. Few questions...

In science it is always a challenge to derive a general law from specific empirical observations. Those who take this challenge should be able to answer a few questions regarding the validity of the extrapolation. For TRIZ-based tools these questions are quite obvious, for example:

(i) Is Matrix of Contradictions statistically stable to a number of patents analyzed? Or simply speaking: If we continue to perform analysis of patents similar to that which TRIZ founder Genrikh Altshuller performed, and

APPENDIX B

eventually analyze all existing patents, will Contradiction Matrix change? If it will, then how will it change?

(ii) Are discovered trends of technology evolution statistically stable to a number of patents analyzed? Do they cover all existing trends in the current world of technology?

(iii) How does one cross a chasm from a general recommendation to a specific innovative idea?

To answer these questions, we need some tools capable of reading millions of documents (e.g., patents) and being 'intelligent' enough to 'understand' them. The tool providing such means is based on semantic indexing technology; its application to the problem-solving practice launches a novel approach to the innovation process which eventually may become the new engineering discipline - Semantic TRIZ™.

3. Basics of Semantic Indexing Technology

Semantic Indexing Technology is based on mathematical linguistics. Linguistic analysis of the natural language text [3] is currently performed on four major levels which could be generally defined as sentence and word recognition, lexical analysis, syntactic analysis, and *semantic* analysis. The mission of the first level is obvious. Lexical analysis involves reading the input sentences, extracting individual words, and retrieving the possible word types from the databases (dictionaries). Lexical analysis is enhanced by hidden Markov chains model, which provides probability distribution for word type sequences and determines the most likely sequence of word types in a sentence. Syntactic analysis employs phrase-structure grammars, identifies the syntactic structure of the text and consequently determines the exact word type. Semantic analysis identifies the meaning of the text by extracting from a sentence its semantic items such as subject, action, and object.

Applying this analysis for the following sentence,

Electrolytic dissociation can be successfully used to measure air humidity,

software will determine that in this sentence

The Subject is *electrolytic dissociation*
The Action is *measure*, and
The Object is *air humidity*

These semantic items are of great importance because they contain information about what question can be asked and what answer can be served in response. For example, if someone asks the question: 'How can I measure humidity?', the person who is asking this question in fact defines, that in the possible response Action should be '*measure*' and Object should be '*humidity*'. What is unknown to him is the Subject

APPENDIX B

(*what* measures humidity?). If we performed semantic analysis of all documents available, including one with the above sentence, which means that we would extract all Subjects, Actions, and Objects, then, in response to this question, we would provide only those Subjects, Actions, and Objects combinations which have the Action and Object like in the questions, and this will form an exact answers to the question *(e.g., electrolytic dissociation – measure - air humidity).*

This simple example in fact defines the architecture of the semantic indexing technology which will support Semantic TRIZ™ research:

(i) It should have natural language interface, which is able to extract semantic entities from the question;
(ii) It should have a searchable database of semantic entities extracted from (generally speaking) the entire universe of original documents. The process of creating such a database is called semantic processing and is quite similar to a process of educating a human: we are 'asking' computer to read documents, understand them by means of extracting semantic items, and to memorize this understanding;
(iii) Matching semantic items from a question with semantic items in the database provides exact answers to a question

This architecture was implemented in Goldfire Intelligence™ platform [4], where a searchable database of semantic entities was created by applying linguistic analysis to the entire worldwide collection of patents. In the following chapters, we will demonstrate that this platform is very capable of supporting TRIZ innovation process.

4. Goldfire Intelligence™ as 'Effects-on-Demand' answering engine platform

The most straightforward application of Goldfire Intelligence™ is to ask it direct natural language questions. Figure 4 below illustrates this process.

APPENDIX B

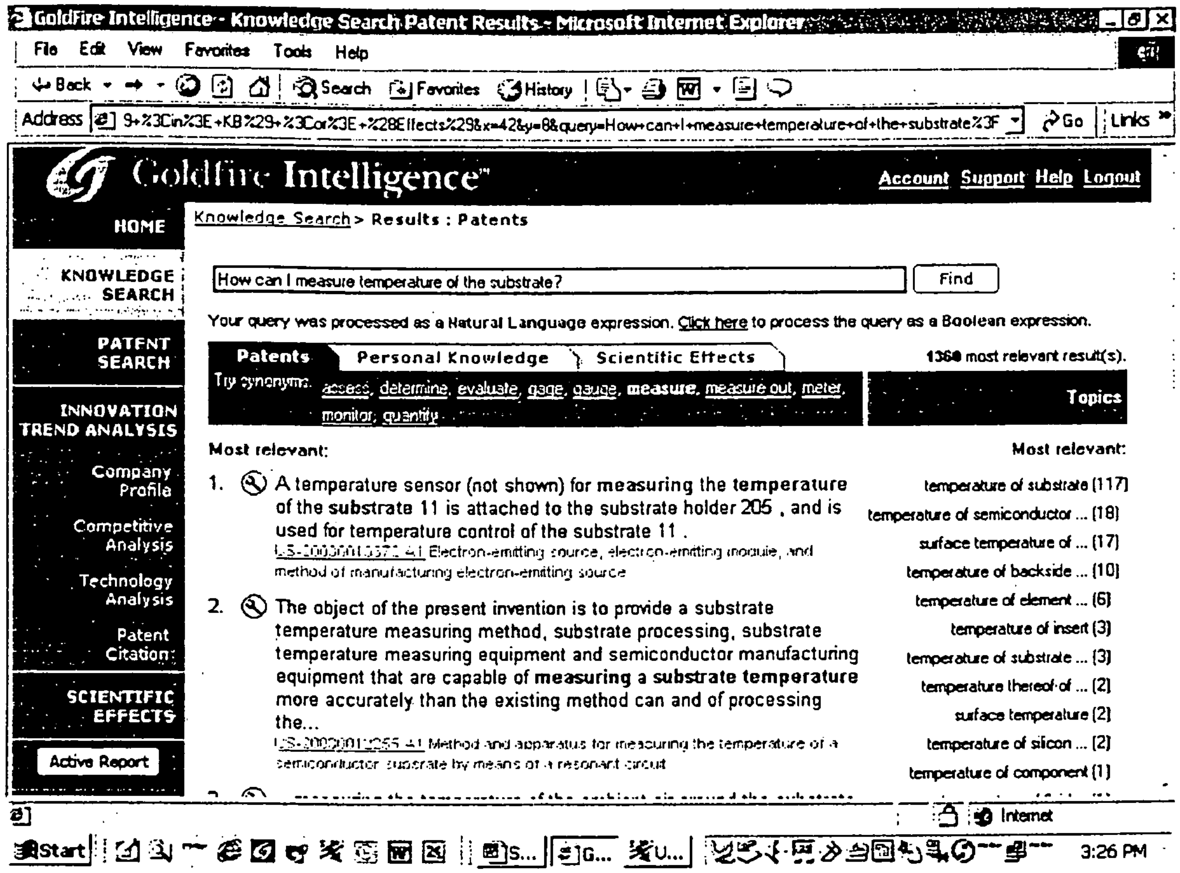


Figure 4: Goldfire Intelligence™ as 'Effects-on-Demand' answering engine platform.

When the question 'How can I measure temperature of the substrate?' is asked, the system performs analysis of the question, determines what semantic items it contains (action – measure, object – temperature of the substrate), and matches it against hundreds of millions of semantic items, extracted from the worldwide collection of patents. The results of the match represent the exact answer to the question and are shown in Figure 4. While the results displayed in Figure 3 demonstrated what physics in general is available to measure temperature, the results of Figure 4 are very specific, indicating how the *substrate* temperature is measured. This is why we can call this application 'customized Effects' or 'Effects-on-Demand'.

5. Goldfire Intelligence™ as Matrix of Contradictions

Questions in the contradiction template can be also addressed by Goldfire Intelligence™. In Figure 1, we illustrated how Altshuller's Contradiction Matrix handles the situation, described by the following statement: 'I want to improve thermodynamic properties of my design by increasing its cross-section area, but there are undesirable consequences of the area increase – volume increases as

APPENDIX B

well'. We can now ask this question as a natural language question: 'How can I increase area, but decrease volume?' and address it again to millions of patents.

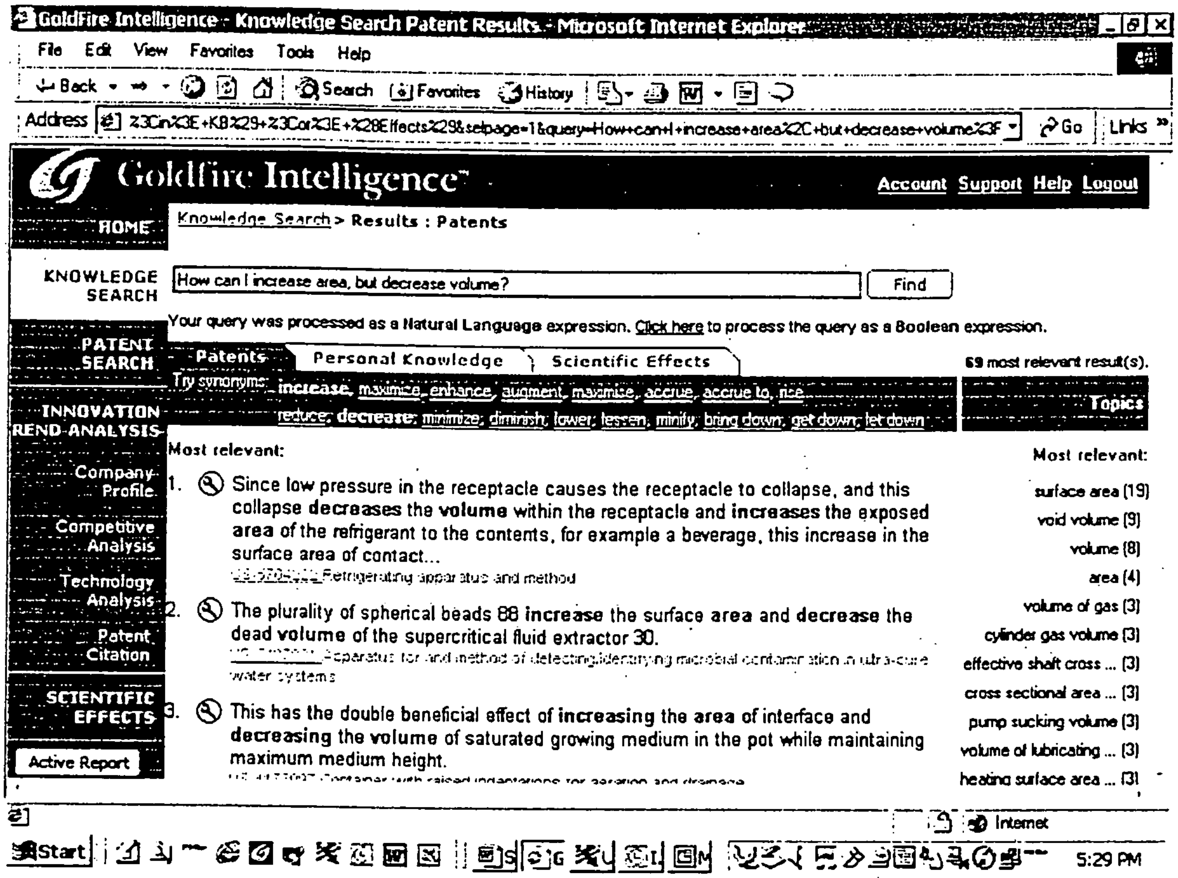


Figure 5: Goldfire Intelligence™ as matrix of contradictions.

The system response is illustrated in Figure 5; it presents very explicit information about how this exact contradiction has been previously solved.

6. Trends of evolution research with Goldfire Intelligence™

Goldfire Intelligence™ enables researchers to investigate trends of evolution for any industry, any technology, any product, design, material, or, generally speaking, it can build a time dependence of answers to any natural language question. This process is presented in Figure 6. It shows that, for example, question 'How can we detect a gas leak?' can be asked in a specific time domain. Asking this question recurrently, we will see how answer to this question evolves in time. Answer-to-question (or solution-to-problem) time dependence is nothing else but specific technology evolution trend. Results for the question 'How can we detect a gas leak?' asked in 5-year intervals are shown in Figure 7.

APPENDIX B

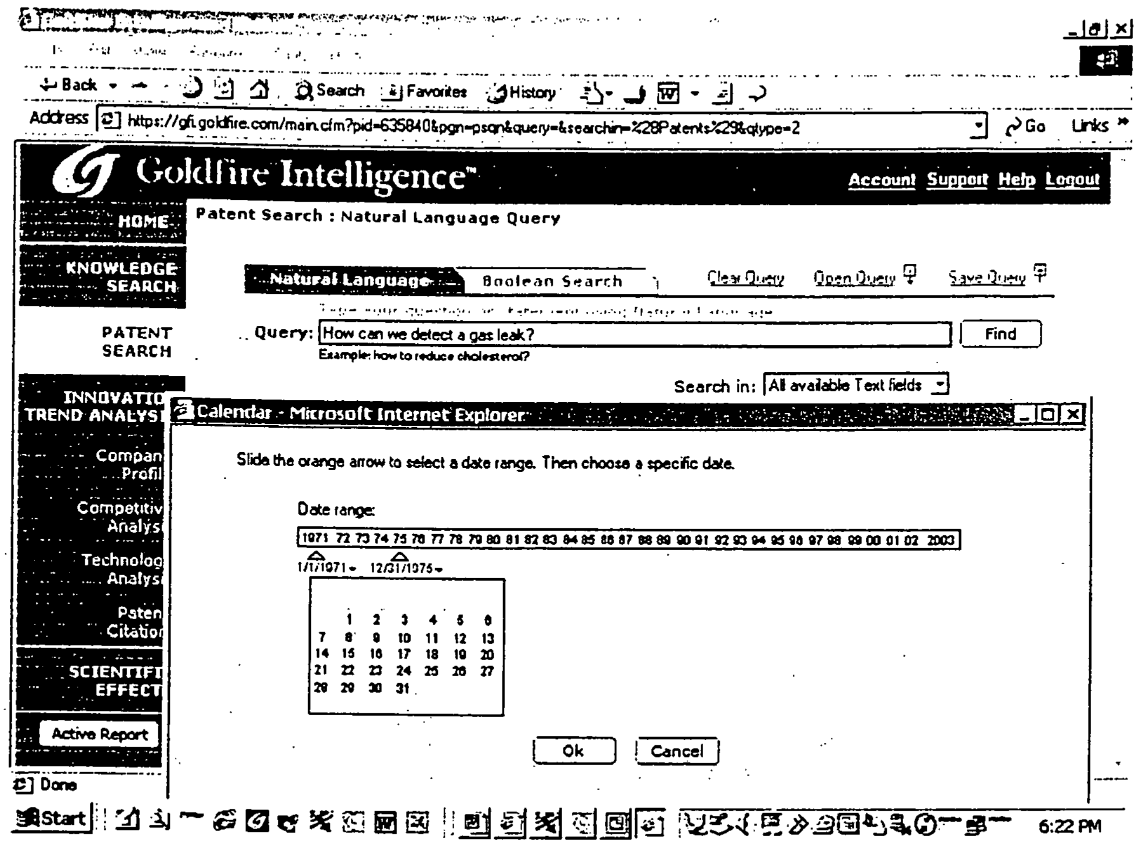


Figure 6: Asking questions in a specific time domain.

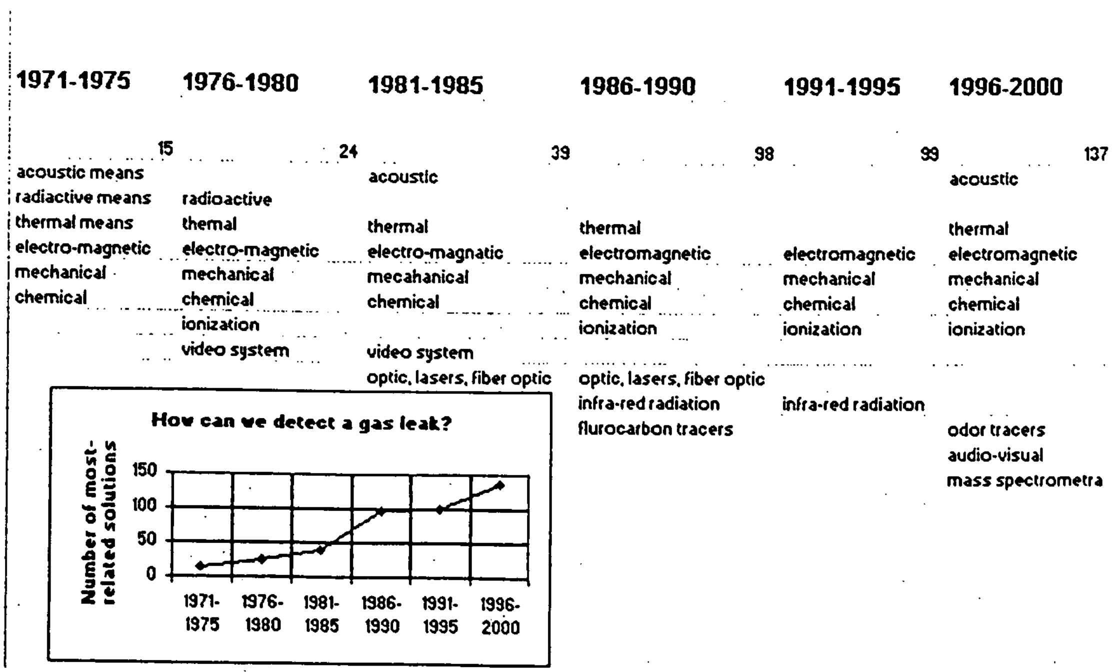


Figure 7: Time dependence of Goldfire Intelligence™ generated answers.

APPENDIX B

It can be clearly determined that gas-leak-detection systems evolved from acoustic, thermal, and mechanical designs to optical, audio-visual, and spectrographic means.

In TRIZ, it is widely believed that the assessment of the system development stage ('infancy' – fast development – maturity S-curves) can be determined when time dependence of the quantity of patents is compared against their innovation level. While number-of-patents versus time functions can be easily calculated, *automatic* innovation level evaluation is very challenging. Five levels of innovation, suggested by Altshuller, are logical, but do not provide the exact criteria for practical usage and therefore are extremely difficult to quantify. To address this problem, we propose to employ patent forward citation as an indicator (certainly, not the unique one) of its level of innovation, speculating that the break-through inventions will be extensively referenced by the followers.

The following statistical model has been adopted. Suppose we have a sequence of related (i.e., same technology) patents $P_1; P_2;...P_i; P_{i+1};...P_n$, where index i increases in time. Then the probability that a single reference in P_{i+1} is directed to the patent P_i can be estimated as

(1) $c = \frac{1}{i}$

The probability that a single reference in P_{i+1} is not directed to the patent P_i can be estimated as

(2) $f = 1 - \frac{1}{i}$

The probability that patent P_i is not referenced at least once by patent P_{i+1}, if patent P_{i+1} has m backward references in total, can be estimated as

(3) $F = \left(1 - \frac{1}{i}\right)^m$

The number of patents K, which being combined would definitely provide at least one reference to the patent P_i, can be estimated from the equation (4):

(4) $\prod_{k=0}^{k=K} F_k = \left(1 - \frac{1}{i}\right)^{m_{i+1}} \left(1 - \frac{1}{i+1}\right)^{m_{i+2}} .. \left(1 - \frac{1}{i+k}\right)^{m_{i+1+k}} .. \left(1 - \frac{1}{i+K}\right)^{m_{i+1+K}} = 0;$

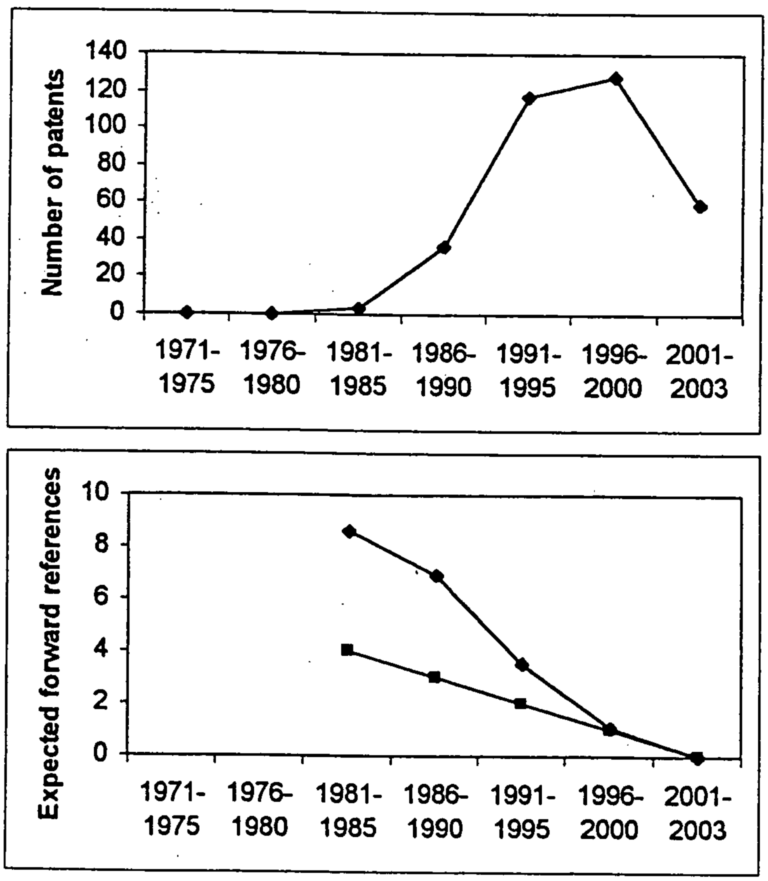
Then, normally expected number of forward references received by patent P_i can be estimated as:

(5) $C \approx \frac{n-i}{1+K}$

APPENDIX B

We will consider patent as being an 'outstanding' one if number of its actual forward references exceeds *C*.

As an example, on Figure 8 we show results of the analysis conducted with Goldfire Intelligence™ for 344 patents mentioning 'fiber optic gyroscope' in their titles or abstracts. Interestingly, we can see that number of patents evolve in a very 'classic' manner clearly exhibiting 'infancy' – fast development – maturity cycle. As it could be also anticipated (newer patents have less chance to be referenced), number of expected forward references decreases in time. Results also demonstrate that number of 'outstanding' patents and especially their share within total number of patents, *may* exhibit the trend when earlier patents are of higher quality, though statistical uncertainty cannot conclusively support this.



APPENDIX B

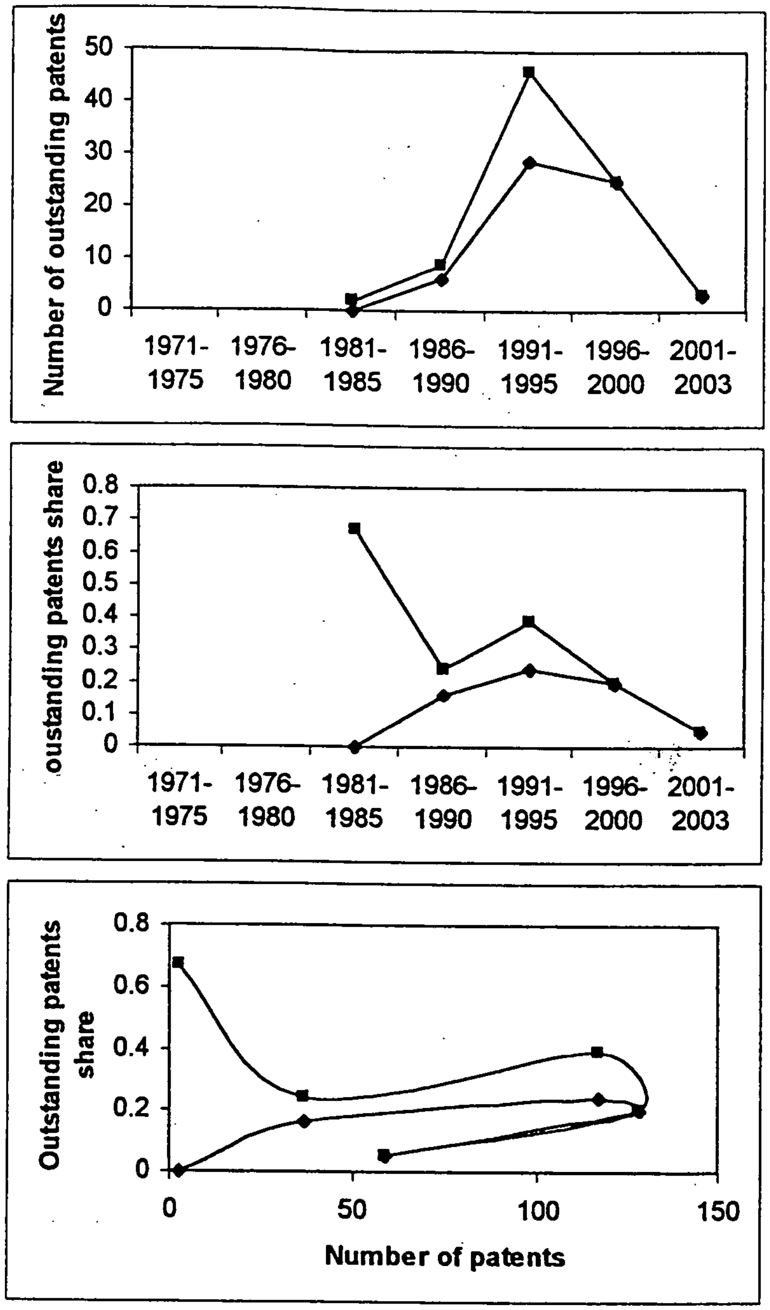


Figure 8: Goldfire Intelligence™ generated evolution of fiber optic gyroscope technology. Blue and pink curves define result's sensitivity to parameters variations.

APPENDIX B

7. Conclusions

Half a century ago, when TRIZ methodology was originated, the ability to automatically study millions of documents was absolutely inconceivable. Therefore traditional TRIZ repeated steps of many empirically derived methods: from extensive (though limited) observations to extrapolation and generalization.

Recent advances of computational linguistics allowed us to combine the classic TRIZ approach to the innovative problem solving process with the benefits of Semantic Indexing Technology. We call this enhanced process Semantic TRIZ™. By its nature, since every question is addressed to the entire worldwide collection of patents, Semantic TRIZ™ is very specific and therefore it can support many traditional TRIZ tools:

(i) If questions are formulated directly, Semantic TRIZ™ works as a customization of scientific effects data base;

(ii) If questions are formulated as a contradiction, Semantic TRIZ™ works as a huge (currently 10^7x10^7) matrix of contradictions providing specific answers on how this contradiction has been solved

(iii) If questions are formulated relative to a specific time domain, Semantic TRIZ™ generates exact trends of technology evolution

Acknowledgement: I am grateful to Jim Todhunter for constructive discussions.

References

1. Ena Arel, Ruth Bowers. *TechOptimizer 4.0 User Guide.* 2002. Invention Machine Corporation, Boston, MA, 374 pp.

2. Ena Arel, Mikhail Verbitsky, Igor Devoino, Sergei Ikovenko. *TechOptimizer Fundamentals*, 2002, Invention Machine Corporation, Boston, MA, 134 pp.

3. *Precision in Knowledge Retrieval.* 2002. Invention Machine Corporation, Boston, MA, 5 pp.

4. Goldfire Intelligence™: http://www.invention-machine.com/prodserv/GFI.cfm

APPENDIX B